



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : G06F 19/00, G01N 27/447	A1	(11) International Publication Number: WO 97/22076 (43) International Publication Date: 19 June 1997 (19.06.97)
<p>(21) International Application Number: PCT/CA96/00834</p> <p>(22) International Filing Date: 12 December 1996 (12.12.96)</p> <p>(30) Priority Data: 08/570,994 12 December 1995 (12.12.95) US</p> <p>(71) Applicant (for all designated States except US): VISIBLE GENETICS INC. [CA/CA]; Suite 100, 700 Bay Street, Toronto, Ontario M5G 1Z6 (CA).</p> <p>(72) Inventors; and (75) Inventors/Applicants (for US only): STEVENS, John, K. [CA/CA]; 540 Huron Street, Toronto, Ontario M5R 2R7 (CA). DUNN, James, M. [CA/CA]; 117 Citadel Drive, Scarborough, Ontario M1K 4S8 (CA). DEE, Gregory [CA/CA]; 1400 Davenport Road, Toronto, Ontario M6H 2G9 (CA). CASSIDY, James, W. [CA/CA]; 458 Thorndale Drive, Waterloo, Ontario N2T 7V4 (CA).</p> <p>(74) Agent: McMAHON, Eileen; Deeth Williams Wall, Suite 400, 150 York Street, Toronto, Ontario M5H 3S5 (CA).</p>	<p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, VZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p>Published With international search report. With amended claims.</p>	
<p>(54) Title: VIRTUAL DNA SEQUENCER</p> <p>(57) Abstract</p> <p>A virtual DNA sequencer combines a plurality of individual DNA sequencers. Samples of DNA or other nucleic acid from subjects are prepared and allocated in real time to particular lanes or sets of lanes in electrophoresis plates of the individual sequencers, with records kept of the allocations. The data resulting from the electrophoresis runs is collected and collated according to the identities of the subjects. The individual sequencers are networked, and each individual sequencer is preferably equipped with a data buffer large enough to accommodate all or substantially all of a data run, thus protecting the virtual sequencer from loss of valuable data in the event that the network is disrupted for some portion of the time of the data run. In this way, a plurality of sequencers is virtually the same as a single sequencer with a very large number of tracks each of which can run for a much longer sequencing run than an individual sequencer.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NI	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	TM	Turkmenistan
EU	Spain	ML	Mali	UG	Uganda
		MT	Malta		
		MU	Mauritius		
		NA	Namibia		
		NC	New Caledonia		
		NE	Niger		
		NG	Nigeria		
		NI	Nicaragua		
		NL	Netherlands		
		NO	Norway		
		NZ	New Zealand		
		OM	Oman		
		PA	Panama		
		PE	Peru		
		PG	Papua New Guinea		
		PH	Philippines		
		PK	Pakistan		
		PL	Poland		
		PT	Portugal		
		RU	Russian Federation		
		RO	Romania		
		RS	Serbia		
		RU	Russian Federation		
		SA	Saudi Arabia		
		SC	Seychelles		
		SD	Sudan		
		SE	Sweden		
		SG	Singapore		
		SI	Slovenia		
		SK	Slovakia		
		SN	Senegal		
		SZ	Swaziland		
		TD	Chad		
		TG	Togo		
		TJ	Tajikistan		
		TM	Turkmenistan		
		TT	Trinidad and Tobago		
		UA	Ukraine		
		UG	Uganda		
		US	United States of America		
		UY	Uruguay		
		UZ	Uzbekistan		
		VC	Virgin Islands		
		VE	Venezuela		
		VG	Virgin Islands		
		VI	Virgin Islands		
		VN	Vietnam		
		WS	Samoa		
		YE	Yemen		
		ZA	South Africa		
		ZM	Zambia		
		ZW	Zimbabwe		

Virtual DNA Sequencer

Technical Field

The invention relates generally to DNA sequencers, and relates particularly to a novel approach for uniting a multitude of DNA sequencers into a large virtual DNA sequencer.

5

Background Art

As more and more diseases are linked to genetic abnormalities, the possibility of diagnosis or of predictive screening for such diseases by sequencing an individual's nucleic acids takes on increasing significance. The first analytical methods for sequencing nucleic acids were slow and expensive. It is thus unsurprising that an extraordinary amount of effort has been directed in recent years to the task of sequencing nucleic acid sequences. The general approach is by now well known even to the general public: a sample of nucleic acid (typically DNA) is replicated (typically by PCR), decomposed into fragments, and passed through an electrophoresis gel. The fragments are tagged and detection of the tags permits determining the nucleotides making up the sequence. In a typical sequencer the tags are fluorescent tags, laser light is used to stimulate the tags, and photomultiplier tubes are used to detect the light that is given off by the fluorescent tags. Each sequencer contains a plate which is divided into a plurality of lanes, each one of which carries a sample being sequenced.

25

One of the steps in nucleotide sequence determination of a subject nucleic acid polymer is interpretation of the pattern of oligonucleotide fragments which results from electrophoretic separation of fragments of the subject nucleic acid polymer (the "fragment pattern"). The interpretation of the fragment pattern, colloquially known as "base-calling,"

30

results in determination of the order of four nucleotide bases, A (adenine), C (cytosine), G (guanine) and T (thymine) for DNA or U (uracil) for RNA in the subject nucleic acid polymer.

5 In the earliest method of base-calling, a method which is still commonly employed, the subject nucleic acid polymer is labeled with a radioactive isotope and either Maxam and Gilbert chemical sequencing (Proc. Natl. Acad. Sci. USA, 74: 560-564 (1977)) or Sanger et al. chain termination sequencing
10 (Proc. Natl. Acad. Sci. USA 74: 5463-5467 (1977)) is performed. The resulting four samples of nucleic acid fragments (terminating in A, C, G, or T(U) respectively in the Sanger et al. method) are loaded into separate loading sites at the top end of an electrophoresis gel. An electric field
15 is applied across the gel, and the fragments migrate through the gel. During this electrophoresis, the gel acts as a separation matrix. The fragments, which in each sample are of an extended series of discrete sizes, separate into bands of discrete species in a channel along the length of the gel.
20 Shorter fragments generally move more quickly than larger fragments. After a suitable separation period, the electrophoresis is stopped. The gel may now be exposed to radiation sensitive film for the generation of an autoradiograph. The pattern of radiation detected on the
25 autoradiograph is a fixed representation of the fragment pattern. A researcher then manually base-calls the order of fragments from the fragment pattern by identifying the step-wise sequence of the order of bands across the four channels.

 More recently, with the advent of the Human Genome
30 Organization and its massive project to sequence the entire human genome, researchers have been turning to automated DNA sequencers to process vast amounts of DNA sequence information. Existing automated DNA sequencers are available from Applied Biosystems, Inc. (Foster City, CA), Pharmacia Biotech,
35 Inc. (Piscataway, NJ), Li-Cor, Inc. (Lincoln, NE), Molecular Dynamics, Inc. (Sunnyvale, CA) and VLSI Technology, Inc.

(Toronto). Automated DNA sequencers are basically electrophoresis apparatuses with detection systems which detect the presence of a detectable molecule as it passes through a detection zone. Each of these apparatus, therefore, are capable of real time detection of migrating bands of oligonucleotide fragments; the fragment patterns consist of a time based record of fluorescence emissions or other detectable signals from each individual electrophoresis channel. They do not require the cumbersome autoradiography methods of the earliest technologies to generate a fragment pattern.

The prior art techniques for computer-assisted base-calling for use in automated DNA sequencers are exemplified by the method of the Pharmacia A.L.F.TM sequencer. Oligonucleotide fragments are labeled with a fluorescent molecule such as fluorescein prior to the sequencing reactions. Sanger et al. sequencing is performed and samples are loaded into the top end of an electrophoresis gel. Under electrophoresis the bands of species separate, and a laser at the bottom end of the gel causes the fragments to fluoresce as they pass through a detection zone. The fragment patterns are a record of fluorescence emissions from each channel. In general, each fragment pattern includes a series of sharp peaks and low, flat plains; the peaks representing the passage of a band of oligonucleotide fragments; the plains representing the absence of such bands.

To perform computer-assisted base-calling, the A.L.F. system executes at least four discrete functions: 1) it smooths the raw data with a band-pass frequency filter; 2) it identifies successive maxima in each data stream; 3) it aligns the smoothed data from each of the four channels into an aligned data stream; and 4) it determines the order of the successive maxima with respect to the aligned data stream.

Despite years of effort, sequencers continue to be slower and more expensive than would be desired.

In the pure research setting it is perhaps acceptable if

5 a sequencer is slow and expensive, and if it is awkward to manage the data yielded by the sequencer. But when it is desired to use DNA sequencers in a clinical diagnostic setting, several further problems arise. A first problem is that each DNA sequencer is treated as an isolated individual system and data must be transferred via a floppy disk or over a network connection to a central computer.

10 The sheer size of the human genome makes it unrealistic to sequence the entirety of the DNA of a patient, so instead it is commonplace to sequence a number of exons, portions of genetic material that are thought to be of interest. This leads to a bookkeeping or paperwork task. In a clinical diagnostic setting, where several exons are to be sequenced for each of thousands of patients, a technician has to keep track of information regarding which exon was run in which lane of which gel plate. Analysis of the resulting data requires transferring files from one floppy disk to another so that an individual exon may be compared with another exon. This is shown schematically in Fig. 1, where the outputs of several sequencers 20a, 20b, etc. are received on floppy disks 21a, 21b, etc. The disks are distributed manually through what is colloquially termed a "sneaker net" or "Nike™ net" (named after the well-known brand of athletic shoes). A series of manual and error-prone steps can lead to data being stored according to particular patient folders 23a, 23b. Similarly if exons are to be compared, this too requires manual steps via the "sneaker net".

30 As suggested in prior art Fig. 2, some prior-art sequencer systems permit connection of sequencers to each other, and to one or more hosts, via a network 24. Or, more commonly, each sequencer 20a, 20b is connected by a dedicated serial link to a personal computer, and the personal computers are connected by a network. In any of these cases, some computer 27 eventually comes to hold files 25a, 25b etc. which represent data collected in the sequencing activities of the sequencers. If an individual exon is to be compared, or for

35

results relating to a particular patient task to be stored together, time-consuming and error-prone manual manipulations 26 are required. The "sneaker network" of Fig. 1 is replaced in part by a network 24, but the time-consuming and error-prone manual manipulations remain.

A further difficulty in a clinical diagnostic setting is that it is awkward or impossible to assign separate tests to separate lanes for different patients. Capacity of the gels is compromised and as a result, often a gel is run with empty lanes.

Analyzing data from different exons from the same patient is very difficult, because in most cases the exons were run on different DNA sequencers on different days, and have been stored in many different disks. Comparing the same exon from many different patients is also difficult for the same reasons.

To be sure the PCR chemistry worked properly, it is often necessary to run the PCR products on a gel. In general such runs could be run on the same gels used for DNA sequencing. Because it is difficult to keep track of results, these are often run as separate batches on one DNA sequencer. This also can waste gel lane space.

Disclosure of Invention

In keeping with the invention, a virtual DNA sequencer combines a plurality of individual DNA sequencers. Samples of DNA or other nucleic acid from subjects are prepared and allocated in real time to particular lanes or sets of lanes in electrophoresis plates of the individual sequencers, with records kept of the allocations. The data resulting from the electrophoresis runs is collected and collated according to the identities of the subjects. The individual sequencers are networked, and each individual sequencer is preferably equipped with a data buffer large enough to accommodate all or substantially all of a data run, thus protecting the virtual

sequencer from loss of valuable data in the event that the network is disrupted for some portion of the time of the data run. In this way, a plurality of sequencers is virtually the same as a single sequencer with a very large number of tracks each of which can run for a much longer sequencing run than an individual sequencer.

Brief Description of Drawings

The drawing will be described in connection with a drawing in several figures, of which:

Fig. 1 is a prior art portrayal of a group of DNA sequencers;
Fig. 2 is a prior art portrayal of a group of DNA sequencers tied together in a network;
Fig. 3 is a portrayal of a group of DNA sequencers comprising a virtual DNA sequencer, according to the invention;
Fig. 4 is a functional block diagram of an individual DNA sequencer forming part of a virtual sequencer;
Fig. 5A shows the protocol-level link between a sequencer and a host computer, and Fig. 5B shows a different protocol-level link between the sequencer and the host computer;
Fig. 6 shows inputs in a sequencer that are serialized into a serial data stream;
Fig. 7 shows a protocol-level link between a sequencer and a host computer, including two signaling channels;
Fig. 8 shows in dataflow form the data propagation relating to two subjects, their DNA, and the analysis and data storage for the subjects;
Fig. 9 shows in functional block diagram for an individual DNA sequencer forming part of a virtual sequencer including a data buffer; and
Fig. 10 depicts data elements relating to the virtual sequencer of the invention and relations therebetween.

Where possible, like items have been shown with like reference numerals.

Modes For Carrying Out The Invention

Fig. 3 shows in functional block diagram form an embodiment of the virtual sequencer 130 according to the invention. The virtual sequencer 130 may be thought of as an instrumentality that receives a sample or group of samples of nucleic acid material from a patient A 101a, and in a more or less automatic and reliable way the analytical result of the sequencing work appears in a folder 23a, as suggested by the solid and dotted arrows. It is almost as if there were a single sequencer with the capability of handling all of the sequencing needs for the task of patient 101a in a single data run. Similarly, and simultaneously, the instrumentality receives a sample or group of samples from patient 101b, and that, too, leads to a folder 23b. The virtual sequencer is or may be much bigger than that shown in Fig. 3, accommodating a large number of individual sequencers and patient sequencing tasks at a particular time. The manner in which this virtual sequencer accomplishes its goals is set forth in detail below.

Fig. 4 shows a typical and preferred individual sequencer that makes up a portion of the virtual sequencer according to the invention. As set forth more fully in U.S. appl. no. 08/353,932 which is incorporated by reference and in PCT appl. no. PCT/US95/14531, published as WO 96-13717, May 9, 1996, which is incorporated by reference, the sequencer starts with a laser diode 31. The laser diode 31 emits a laser beam which reflects from mirror 32 and mirror 33 to form an illuminated region within electrophoresis plate 34. The plate 34 may preferably be that set forth in copending US appl. no. 08/332,577 and PCT appl. no. PCT/US95/14531, published as WO 96-13717, incorporated herein by reference.

From time to time, the nucleic acid fragments that propagate through the plate 34 present fluorescence in the detection area of the plate 34 as is well known to those skilled in the art of fluorescent-tagged nucleic acid electrophoresis. The emitted light is received in photodiode

detectors 35. The signal from a detector 35 is converted to digital form in an A/D convertor, preferably in the manner set forth in copending US appl. no. 08/497,202, incorporated herein by reference. The digital data stream is handled by a processor circuit board 36, and is passed through a conventional serial port 37. In the system according to the invention this serial data passes to a protocol convertor 38 which is connected to a network 24.

As described in pending U.S. appl. no. 08/452,719, which is incorporated herein by reference, the sequencer is made up of a plurality of detection areas, each of which receives radiation from an electromagnetic radiation emitter. The number of detection areas is preferably a multiple of 4, such as 16 or 40. Each detection area lies in a "track" running in the direction of the voltage gradient. The electromagnetic radiation generated in the detection area is detected by an electromagnetic radiation detector 35. The detector has an electrical output indicative of detected electromagnetic radiation.

Turning now to Fig. 5A there is shown a sequencer 20 of a type with a serial output port 37. The communications channel from port 37 is a serial data link 46 such as an RS-232 link, to a serial input port 43 of a host, typified by personal computer 27. Software 42 in the computer receives the data stream via the serial port 43. The typical handshake lines request-to-send, clear-to-send, data-terminal-ready and data-set-ready are employed along with a transmit data line and a receive data line.

Turning momentarily to Fig. 6 there is shown in input-output form a typical data flow in the sequencer. This sequencer has sixteen tracks and data channels 50 which provide a sixteen-bit word to be communicated externally from the sequencer. In addition, it is preferable to provide eight sixteen-bit words of status information 51 in the data stream, indicative of the electrophoresis voltage, plate temperature, and other important information. Thus, each packet of data

contains twenty-four words or forty-eight bytes of data.

In one sequencer system it has been established to provide one packet per second. However, with advances in sequencing speed (as described in the above-referenced
5 copending patent applications) there is the possibility that an event of interest would be missed if data were only collected once per second. Thus, it is considered preferable to provide such a packet every 250 msec, or four packets per second.

10 Returning now to Fig. 5B there is shown a system like that of Fig. 5A, except that a protocol conversion box 38 is connected between the serial port 37 and the network 24. The PC 41 has a network interface card 45. A software object 44 is interposed between the software 42 and the network
15 interface card 45. The result of the software object 44 is that from the point of view of the software 42 the connection is identical to that of Fig. 5A; the software 42 need not be changed despite introduction of the protocol convertor 38.

Turning now to Fig. 7 there is shown a sequencer 20 and
20 host computer 41 in more detail, in a preferred embodiment. The sensors provide analog signals to an A/D convertor 60. The output of the A/D convertor 60 is buffered in a buffer 39. This buffer is selected to be a size comparable to the entire data output of the sequencing run of the sequencer 20. For
25 example, in the case where a 48-byte packet is sent every 250 msec., then the data rate is about 192 bytes per second. If a buffer 39 of 1.2 MB (megabytes) is provided, then about 100 minutes of data may be stored. On the other hand, the buffer might be as little as 1 MB or as much as 2 MB.

30 Stated differently, there is simply a 1.2 MB data storage capacity in each machine. Under normal operation, the data from the sequencer feeds into this memory buffer 39. The data is periodically called for over the network 24 by computer workstations such as host 41, on the network 24. The crucial
35 advantage for this is that if the network 24 fails, then the data is not lost. When the network 24 is deactivated, the

workstation 41 samples the memory buffer 39 to recover any data not yet sent.

This may be compared, for example, with the sequencer of Fig. 5A, which has its own PC 27 attached directly to it.

Such a sequencer has no substantial onboard memory devices. If the PC 27 crashes, then the data run from the individual sequencer 20, prior to PC reactivation, may be lost.

Fortunately, PC's do not crash as often as networks, so problems are relatively.

Fig. 9 shows the sequencer of Fig. 7 in greater detail. With this arrangement, a fully buffered machine 20 can be networked to a much broader network 24, and it is not absolutely required that the network 24 be perfectly reliable. In contrast, many prior-art sequencers have no onboard memory 39. If the system fails to collect the sequencer data when it is ready, the sequencer data is lost. The result can be loss of an expensive data run.

The significance of the relationship between the sequencer 20 and the host personal computer 41 (fig. 7) may be contrasted with that between a computer and, say, a laser printer, which is a sink for data rather than a source.

The extent of the benefit offered by the buffer 29 differs depending on the type of network 24 being used, for example 10Base2 (thin co-axial) and 10BaseT (twisted pair). 10BaseT is cheaper, more resilient, and has a spider topology with a concentrator; and the concentrator rarely goes down. Generally one worries more about a machine crash more than about a network crash. In contrast, where 10Base2 is used, the probability of net crashing is more common because each link must be contained.

Returning to Fig. 7, it was mentioned that the size of the buffer can be anywhere from 1 to 2 MB of data storage. The amount depends on the number of lanes and the sampling rate being used. As mentioned above, the number of lanes may be anywhere in the range of 16 to 40, while the sampling rate may be anywhere from 1 to 32

Fig. 10 depicts pictorially some of the data types that may be used in the system according to the invention. Blocks 111a and 111b represent the raw data from the sequencing machine data runs. Blocks 113a and 113b represent data grouped according to the patient or analysis subject. Block 112 represents laboratory plan data indicating which tracks for a particular sequencer's run are associated with one task (e.g. a section of a patient's DNA) and which are associated with a different task.

It is noted that the data passed from the sequencers to the system are "raw data" from the sensors 35 (fig. 4) and are not nucleotide values. That is, the "base calling" process has not yet occurred at the time the sequencer passes its data to the system. Indeed, it is a preferable aspect of the invention that the raw data are passed in their entirety (or substantially unchanged) to the folders 23a, 23b as suggested in Fig. 8. In this way, even after the patient data have been collected together, it is possible to go back and do the "base calling" again and again in the event that some ambiguity presents itself. Similarly, if it is desired to collect and display several instances of a particular exon (from several different subjects, say) as part of a study or as part of an effort to resolve a base calling ambiguity, the raw data will have been retained permitting such study.

It will thus be appreciated that what has been provided is a virtual sequencer, a system composed of a plurality of individual sequencers. It seamlessly tracks the sub-tasks that make up a sequencing task, permitting the sub-tasks to be split up over several sequencers, and permitting the straightforward collation of the resulting sequence data for study by patient, by task, or by exon. It is as if one had a single sequencer with arbitrarily many electrophoresis tracks, to accommodate an arbitrarily large sequencing task. It is as if one had a sequencer that could sequence far more base pairs in a run than one sequencer can sequence, insofar as the data are seamlessly merged after a sequencing task has been

broken up according to the actual base-pair capacity of the sequencers.

5 Those skilled in the art will have no difficulty devising countless obvious variations of the invention without departing in any way from the invention, all of which are intended to be encompassed by the claims which follow.

Claims

We claim:

1. A virtual nucleic acid analyzer for use in sequencing
respective samples of nucleic acids from first and second
5 subjects, the virtual analyzer comprising:

at least first and second individual nucleic acid analyzers,
each of said individual analyzers comprising at least two
lanes, each of said lanes having a respective analysis region,
each of said individual analyzers further comprising a
10 detector optically coupled with said at least two lanes, said
detector having an output indicative of optical activity in
its respective analysis regions, each of said individual
analyzers further comprising buffer means associated with the
respective detector, said buffer sized to accommodate
15 substantially all the output from an electrophoretic run of
the nucleic acid sample from one of the subjects;

a host;

an input terminal communicatively coupled with the host;

a data store communicatively coupled with the host;

20 a communications channel communicatively coupling the buffer
means associated with the detector of the first and second
individual analyzers with the input terminal of the host;

first means within the host for responding to inputs at the
terminal for storing, within the data store, first records
25 associating the first subject with a first particular lane of
the first individual analyzer and with a first particular lane
of the second individual analyzer, and associating the second
subject with a second particular lane of the first individual
analyzer and with a second particular lane of the second
30 individual analyzer; and

second means within the host for receiving the outputs of the detectors of the first and second individual analyzers and storing, within the data store, second records representative of the outputs and indicative of the lanes of the individual analyzers providing the outputs;

third means within the host for receiving the first and second records, and in response thereto, for storing third records, each of said third records comprising data from particular lanes corresponding to a particular one of the subjects.

2. A nucleic acid analyzer for sequencing a nucleic acid sample, the analyzer comprising:

a plurality of detection areas, each detection area receiving electromagnetic radiation from an electromagnetic radiation emitters, each detection area juxtaposed with an electromagnetic radiation detector, each said electromagnetic radiation detector having an electrical output indicative of detected electromagnetic radiation, the analyzer further comprising buffer means associated with the detector, said buffer sized to accommodate substantially all the output from and electrophoretic run of the nucleic acid sample.

3. The analyzer of claim 1, 2 or 3 wherein each of the buffers is at least one megabyte in size.

4. The analyzer of claim 3 wherein each of the buffers is at least two megabytes in size.

5. The analyzer of claim 2, 3 or 4 further characterized in having at least sixteen lanes.

6. The analyzer of claim 5 further characterized in having at least twenty-four lanes.

7. A virtual nucleic acid analyzer for use in sequencing nucleic acids from first and second subjects, the virtual analyzer comprising:

5 at least first and second individual nucleic acid analyzers, each of said individual analyzers comprising at least two lanes, each of said lanes having a respective analysis region, each of said individual analyzers further comprising a detector optically coupled with said at least two lanes, said
10 detector having an output indicative of optical activity in its respective analysis regions;

a host;

an input terminal communicatively coupled with the host;

a data store communicatively coupled with the host;

15 a communications channel communicatively coupling the outputs of the detectors of the first and second individual analyzers with the input terminal of the host;

20 first means within the host for storing, within the data store, first records associating the first subject with a first particular lane of the first individual analyzer and with a first particular lane of the second individual analyzer, and associating the second subject with a second particular lane of the first individual analyzer and with a second particular lane of the second individual analyzer; and

25 second means within the host for receiving the outputs of the detectors of the first and second individual analyzers and storing, within the data store, second records representative of the outputs and indicative of the lanes of the individual analyzers providing the outputs;

third means within the host for receiving the first and second records, and in response thereto, for storing third records, each of said third records comprising data from particular lanes corresponding to a particular one of the subjects.

5 8. The virtual analyzer of claim 1 or 7 wherein each of the individual analyzers has at least sixteen lanes.

9. The virtual analyzer of claim 8 wherein each of the individual analyzers has at least twenty-four lanes.

10 10. The virtual analyzer of claim 1, 7, 8 or 9 wherein the number of individual analyzers is no fewer than four.

11. The analyzer of claim 1, 7, 8, 9 or 10 wherein the communications channel comprises a network.

12. The virtual analyzer of claim 1, 7, 8, 9, 10 or 11 wherein the input terminal further comprises a bar-code reader and wherein the tracks are labeled with bar codes, the inputs at the input terminal comprising bar code data read by the bar code reader.

13. The virtual analyzer of claim 1, 7, 8, 9, 10, 11 or 12 wherein the first and second subjects represent corresponding analysis subjects from first and second patients.

14. The virtual analyzer of claim 1, 7, 8, 9, 10, 11 or 12 wherein the first and second subjects represent first and second analysis subjects with respect to a particular patient.

15. The analyzer of claim 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 or 14 wherein the detector of each individual analyzer generates its output no less often than once per second.

16. The analyzer of claim 15 wherein the detector of each

individual analyzer generates its output no less often than once per 250 milliseconds.

AMENDED CLAIMS

[received by the international Bureau on 21 April 1997 (21.04.97);
original claims 1-4 amended; remaining claims unchanged (2 pages)]

5 1. A virtual nucleic acid analyzer for use in sequencing
respective samples of nucleic acids from first and second
subjects, the virtual analyzer comprising:

10 at least first and second individual nucleic acid analyzers,
each of said individual analyzers comprising at least two
lanes, each of said lanes having a respective analysis region,
each of said individual analyzers further comprising a
15 detector optically coupled with said at least two lanes, said
detector having an output indicative of optical activity in
its respective analysis regions, each of said individual
analyzers further comprising buffer means associated with the
respective detector, said buffer means sized to accommodate
15 substantially all the output from an electrophoretic run of
the nucleic acid sample from one of the subjects;

a host;

an input terminal communicatively coupled with the host;

a data store communicatively coupled with the host;

20 a communications channel communicatively coupling the buffer
means associated with the detector of the first and second
individual analyzers with the input terminal of the host;

25 first means within the host for responding to inputs at the
terminal for storing, within the data store, first records
associating the first subject with a first particular lane of
the first individual analyzer and with a first particular lane
of the second individual analyzer, and associating the second
subject with a second particular lane of the first individual
analyzer and with a second particular lane of the second
30 individual analyzer; and

second means within the host for receiving the outputs of the detectors of the first and second individual analyzers and storing, within the data store, second records representative of the outputs and indicative of the lanes of the individual
5 analyzers providing the outputs;

third means within the host for receiving the first and second records, and in response thereto, for storing third records, each of said third records comprising data from particular lanes corresponding to a particular one of the subjects.

10 2. A nucleic acid analyzer for sequencing a nucleic acid sample, the analyzer comprising:

a plurality of detection areas, each detection area receiving electromagnetic radiation from an electromagnetic radiation emitters, each detection area juxtaposed with an
15 electromagnetic radiation detector, each said electromagnetic radiation detector having an electrical output indicative of detected electromagnetic radiation, the analyzer further comprising buffer means associated with the detector, said buffer means sized to accommodate substantially all the output
20 from and electrophoretic run of the nucleic acid sample.

3. The analyzer of claim 1, 2 or 3 wherein each of the buffer means is at least one megabyte in size.

4. The analyzer of claim 3 wherein each of the buffer means is at least two megabytes in size.

25 5. The analyzer of claim 2, 3 or 4 further characterized in having at least sixteen lanes.

6. The analyzer of claim 5 further characterized in having at least twenty-four lanes.

1/8

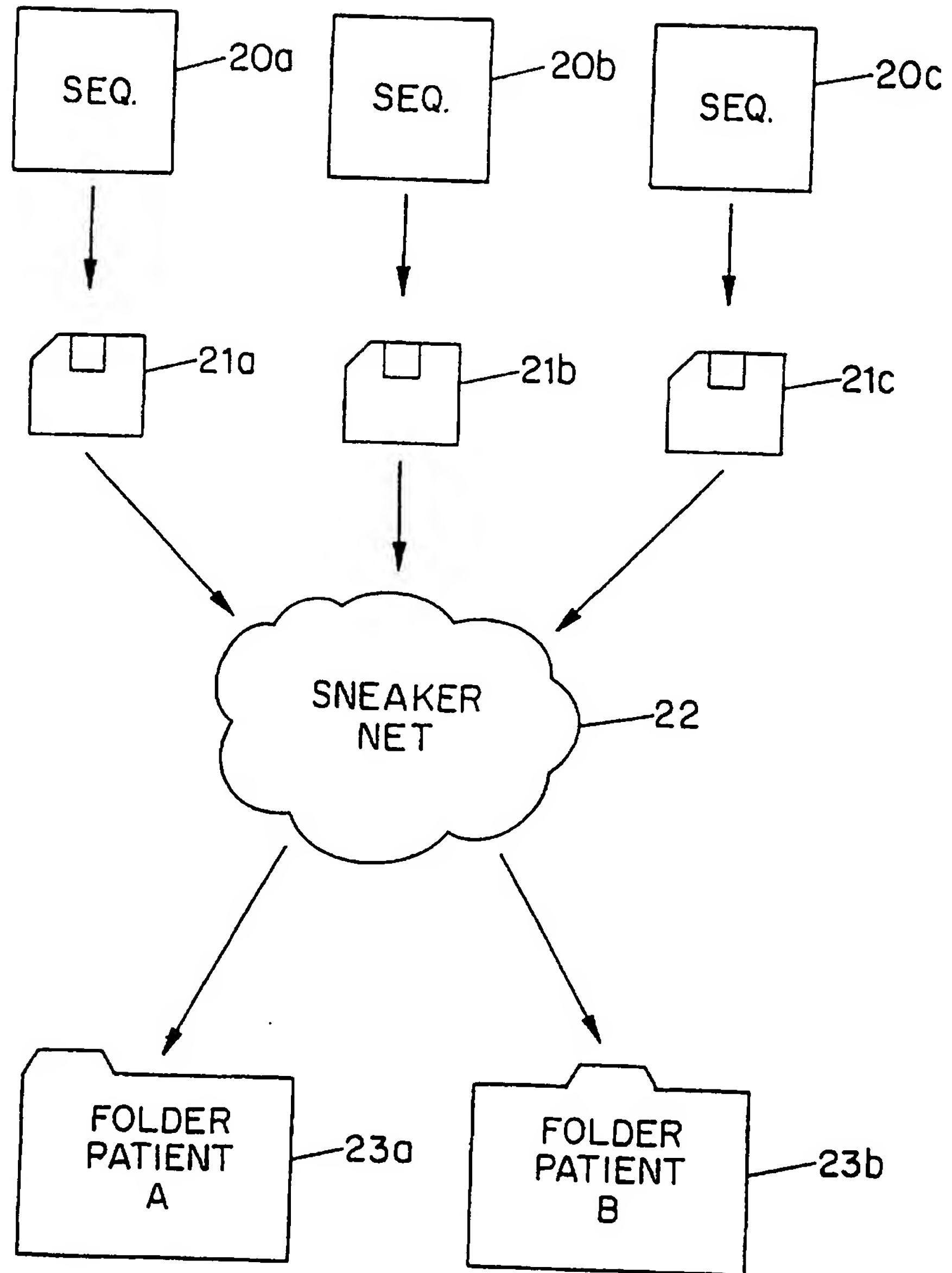


FIG. 1
PRIOR ART

2/8

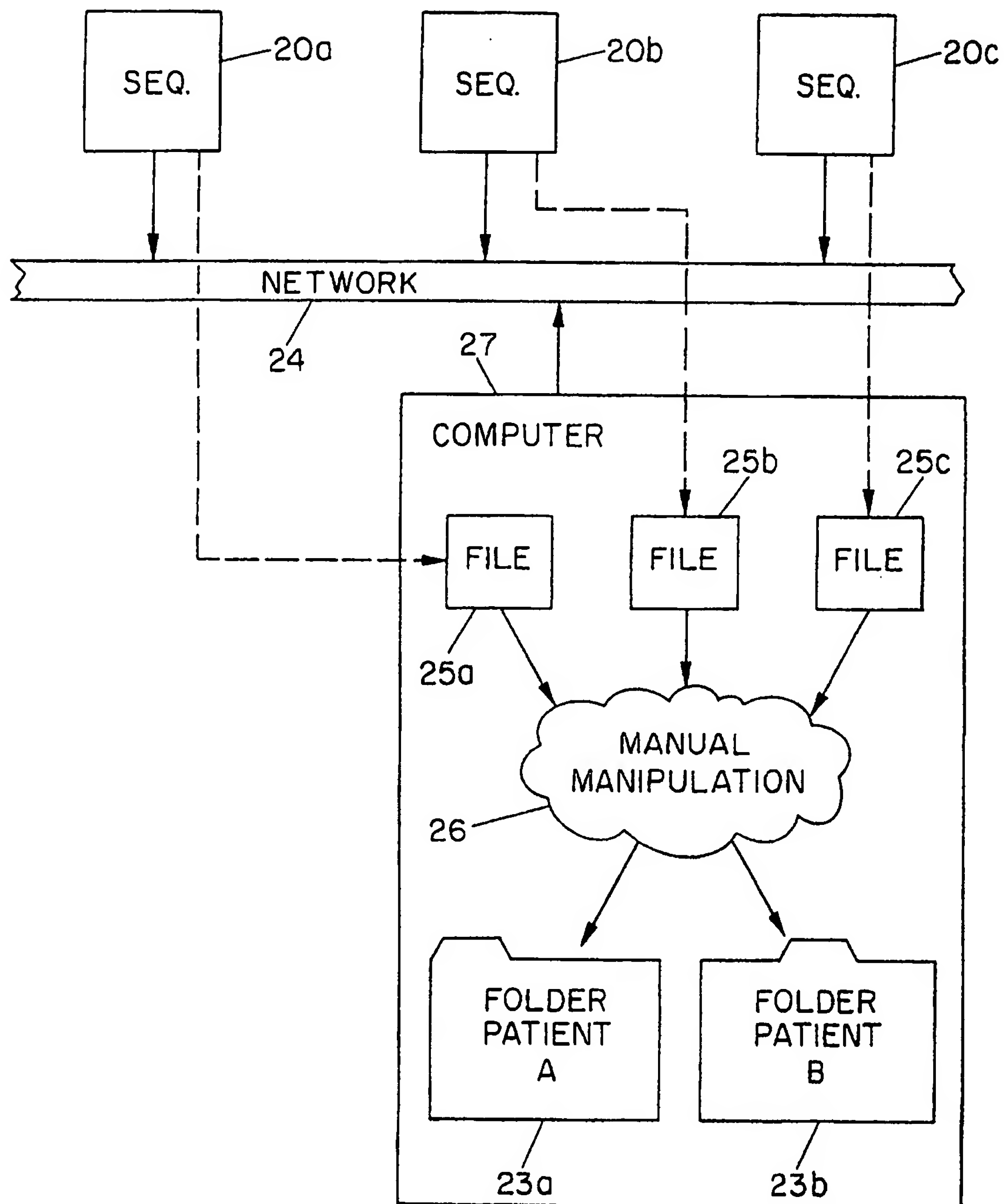


FIG. 2
PRIOR ART

3/8

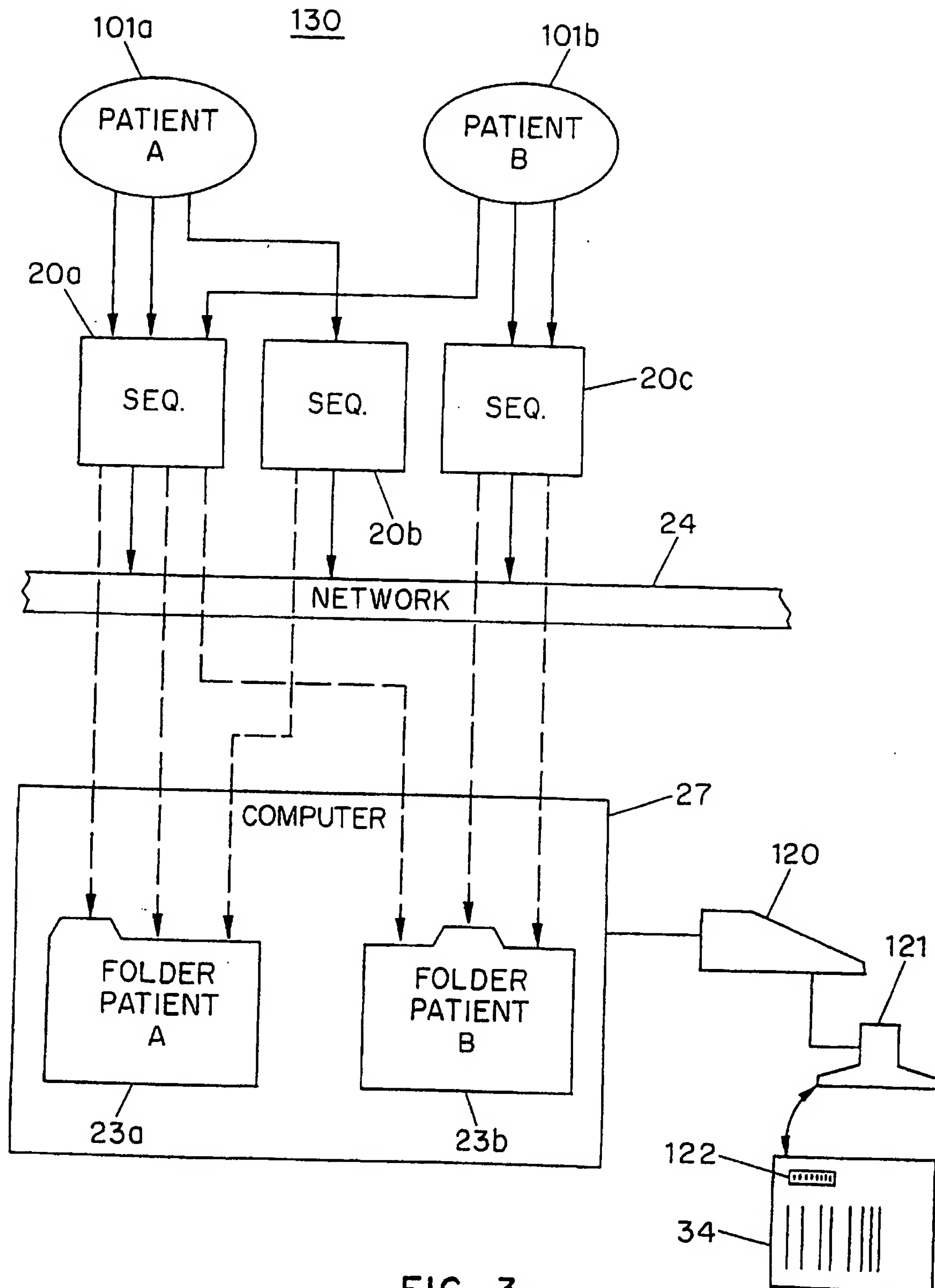


FIG. 3

4/8

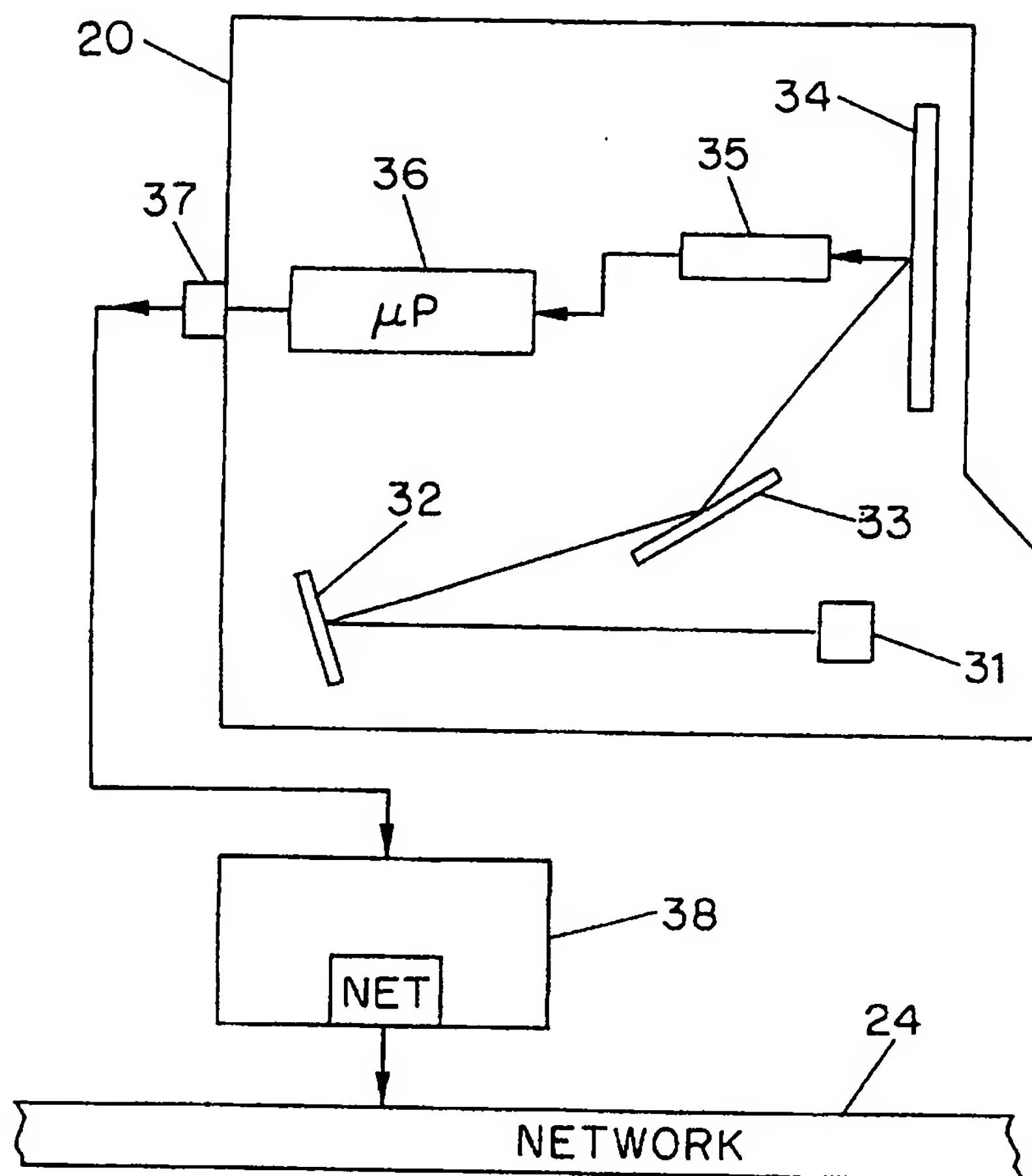


FIG. 4

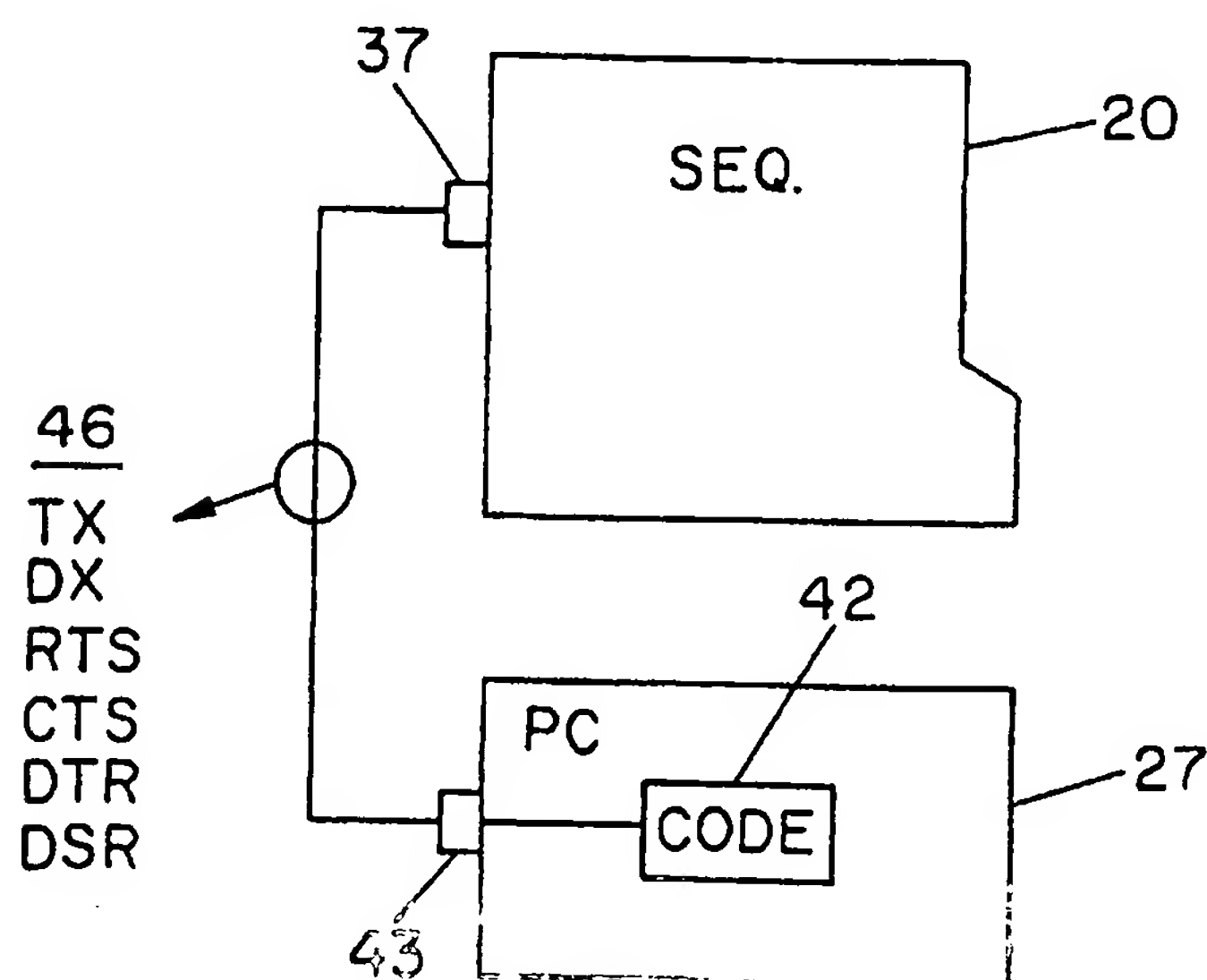


FIG. 5A

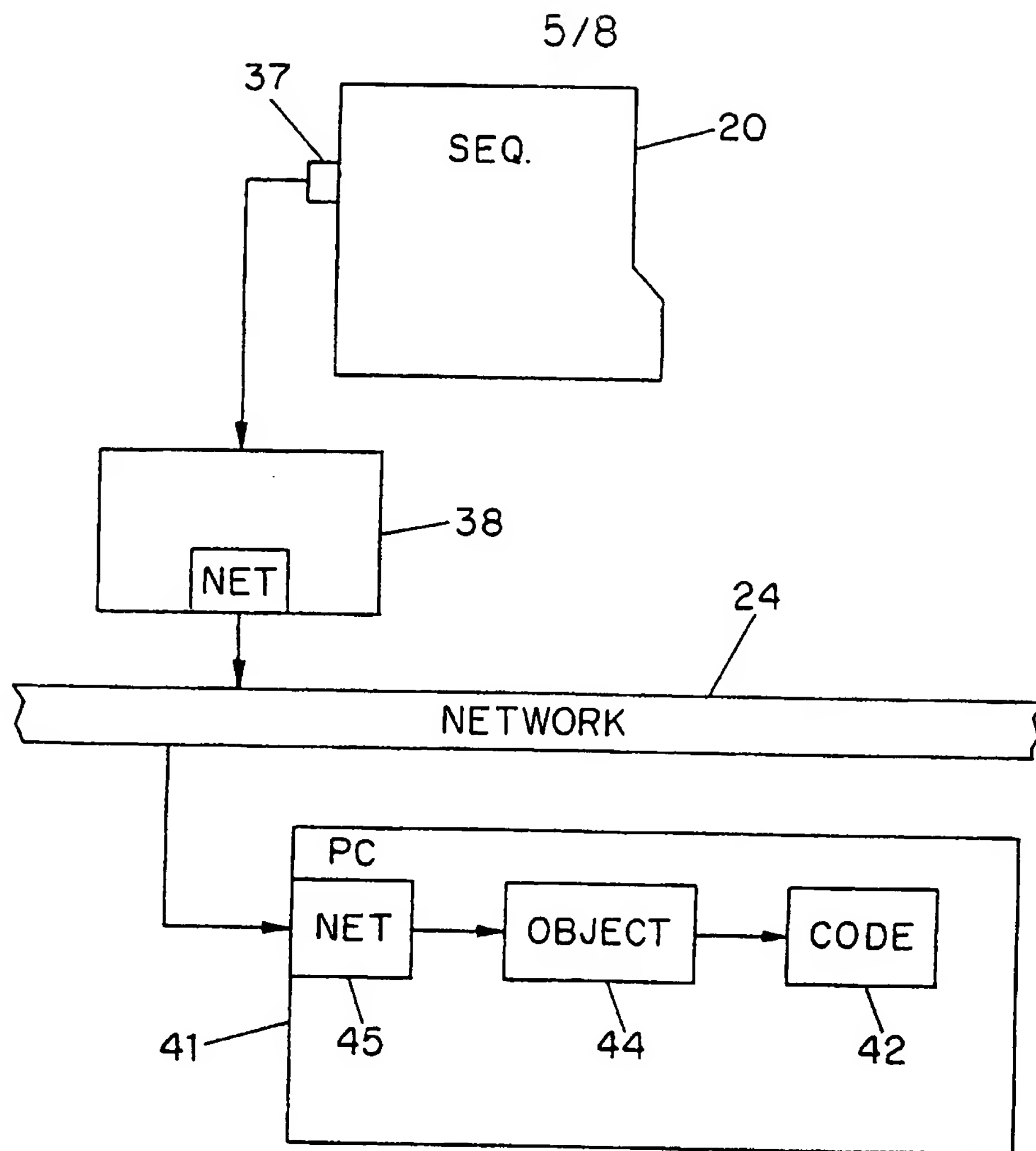


FIG. 5B

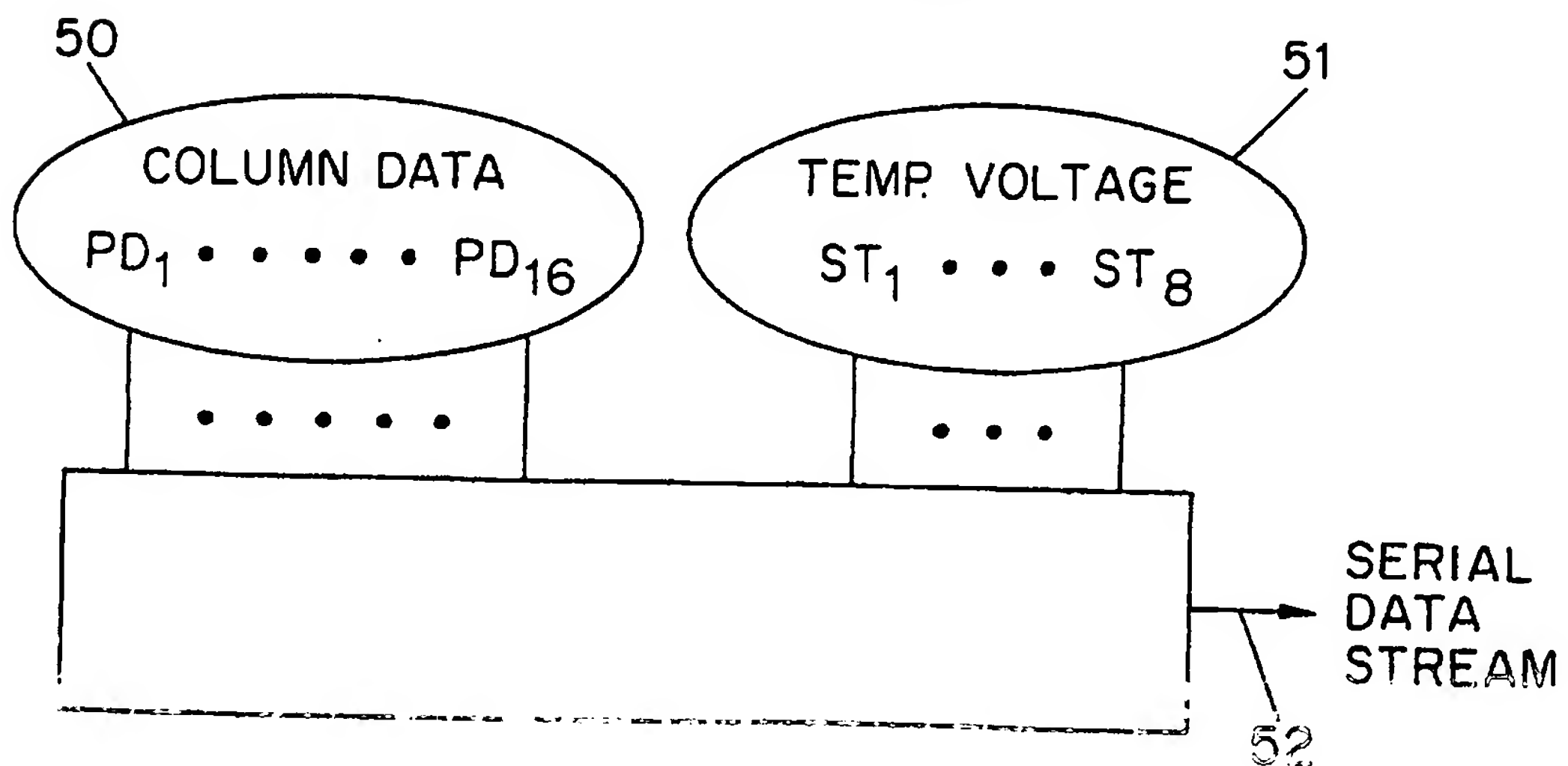


FIG. 6

6/8

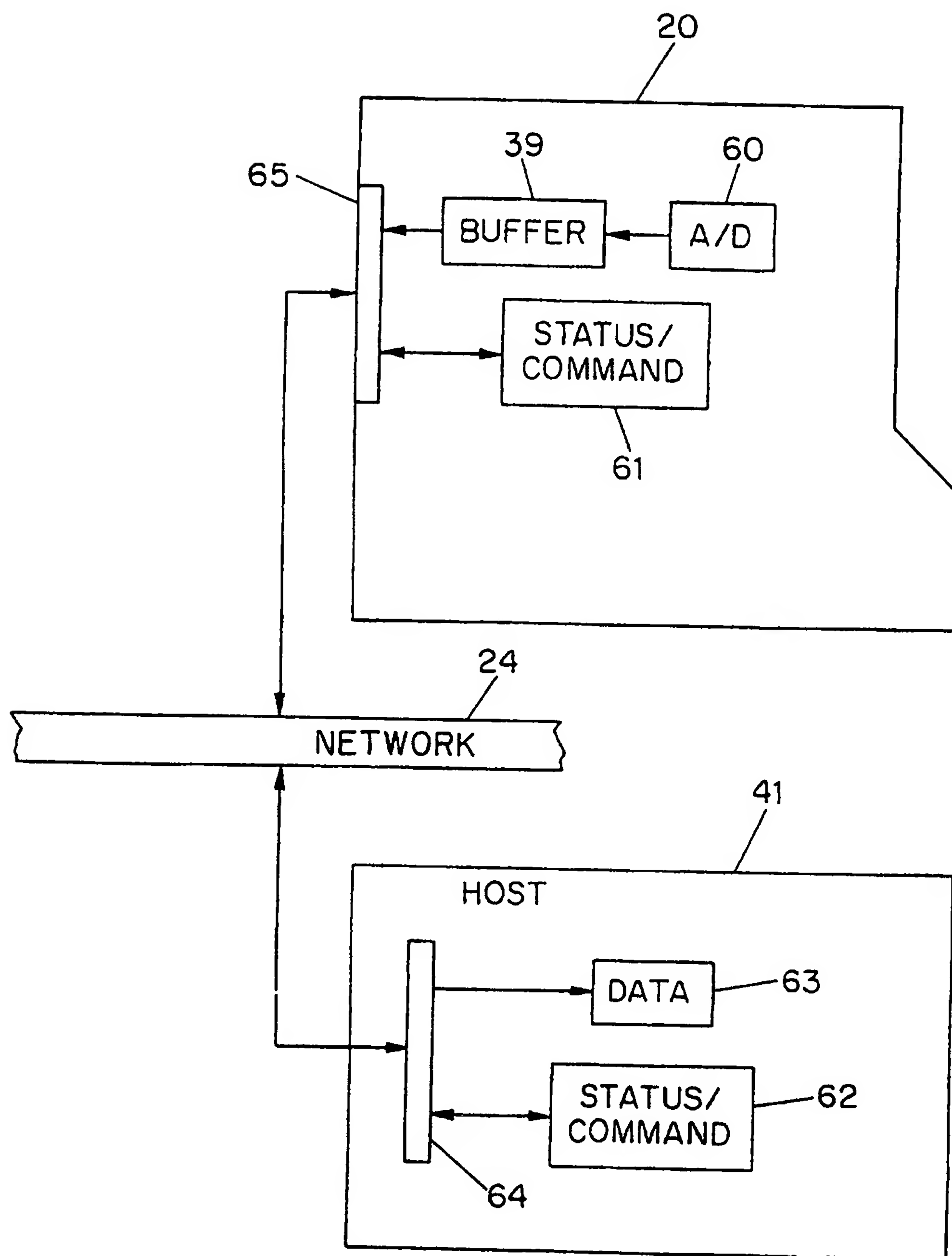


FIG. 7

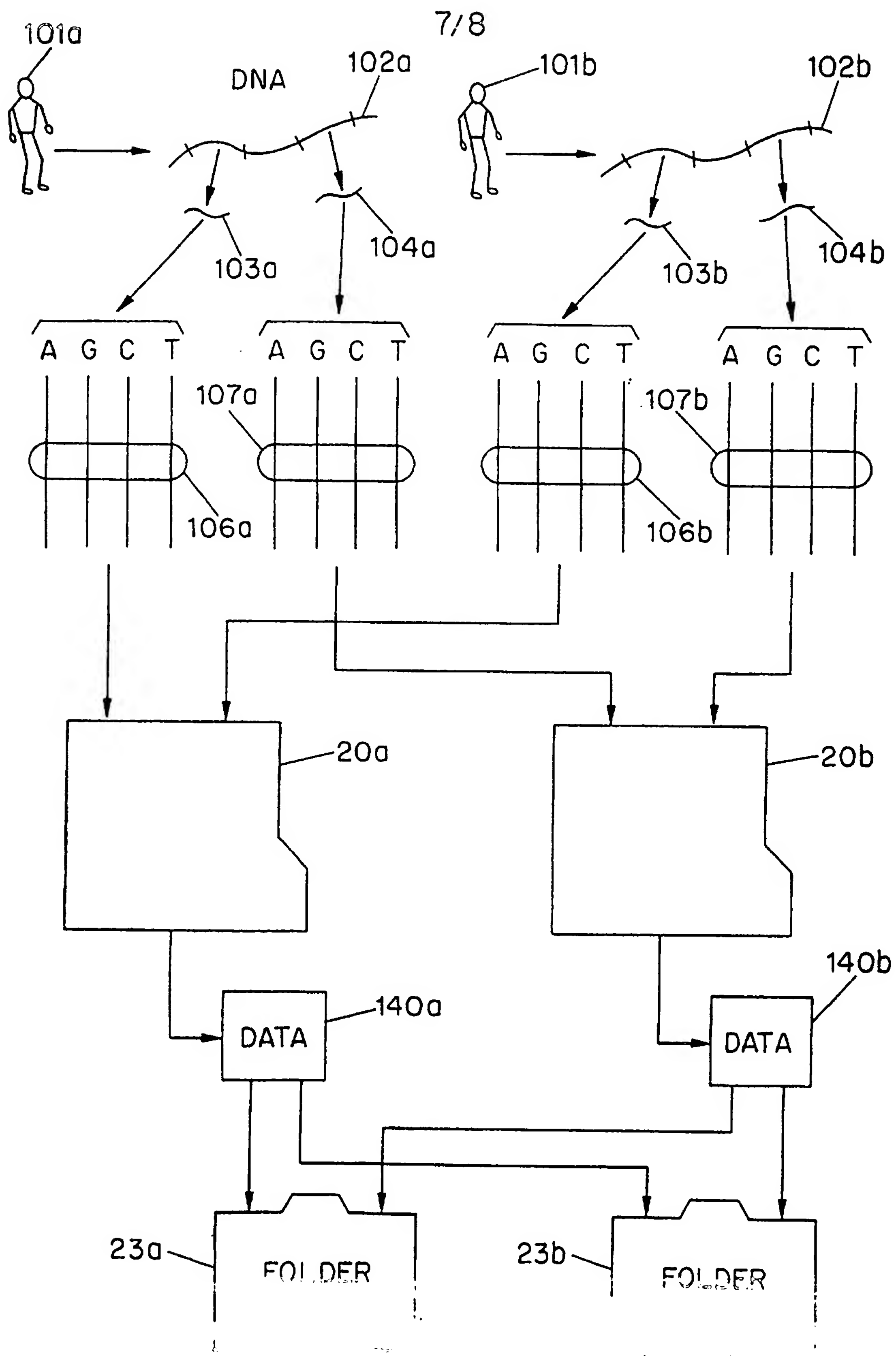


FIG. 8

8/8

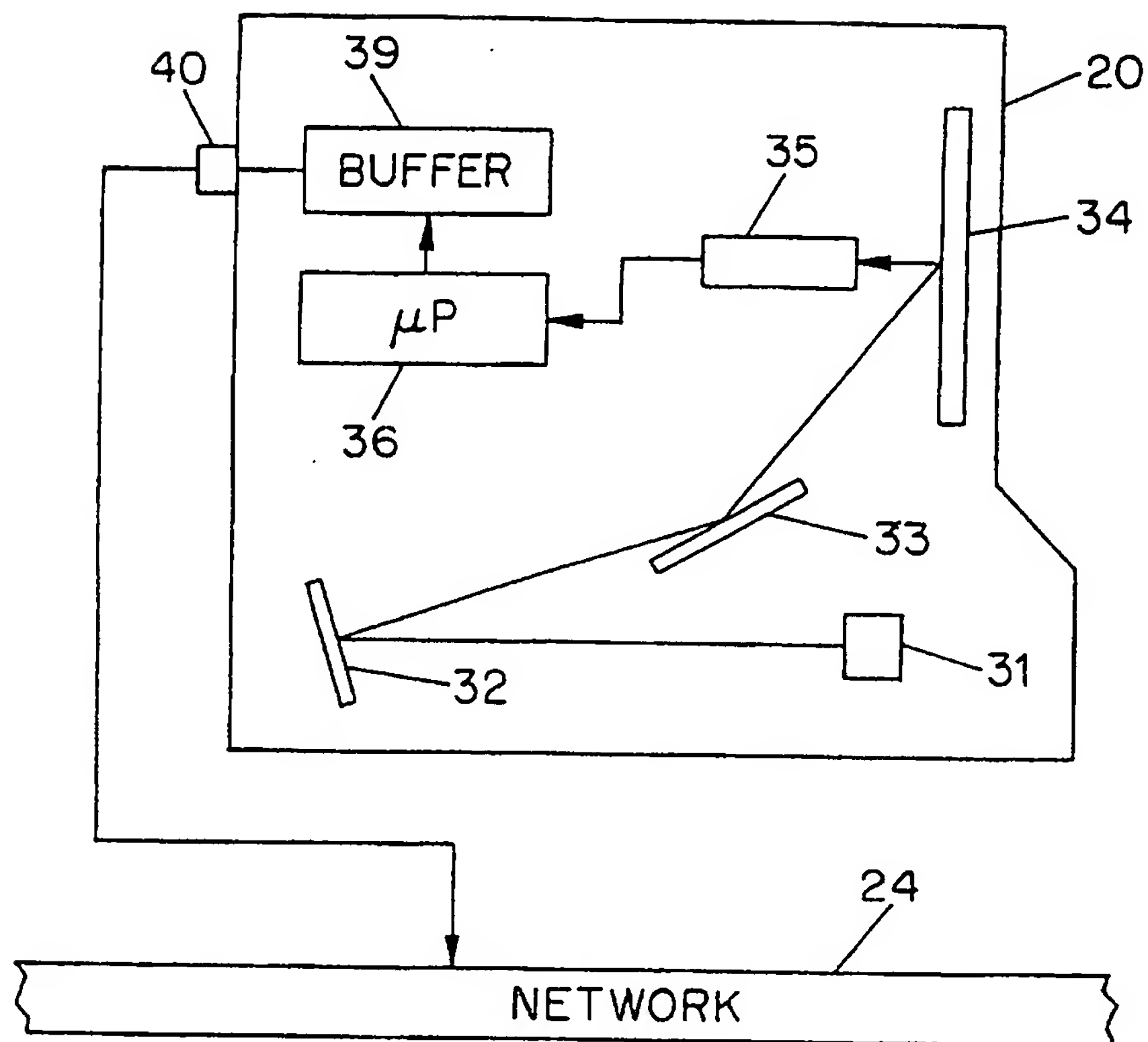


FIG. 9

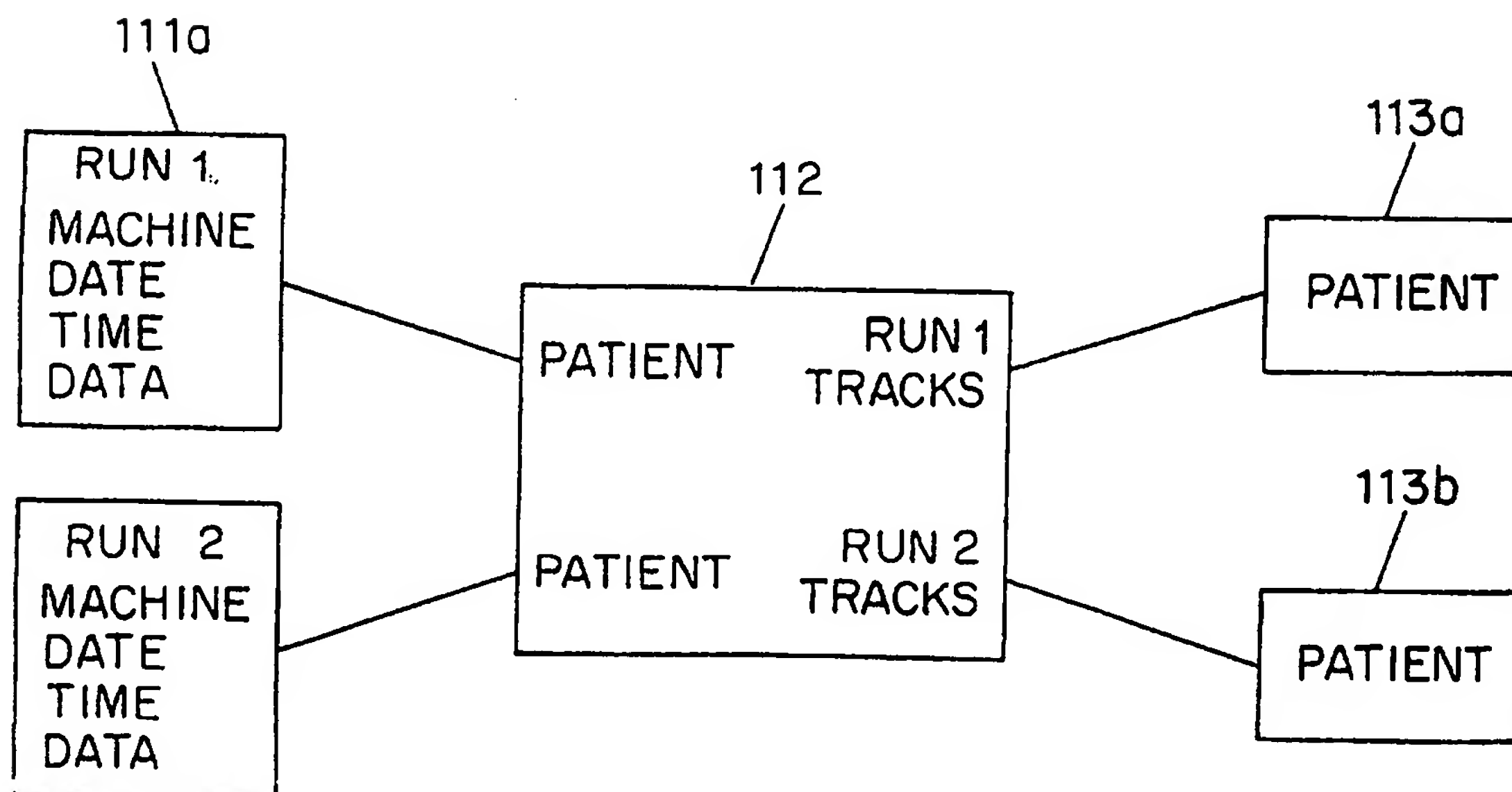


FIG. 10

INTERNATIONAL SEARCH REPORT

International Application No

PC/CA 96/00034

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G06F19/00 G01N27/447

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 G06F G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP,A,0 294 524 (APPLIED BIOSYSTEMS INC) 14 December 1988 see column 3, line 28 - line 53 see column 9, line 42 - column 10, line 3; figure 8 ---	1,2,7
A	EP,A,0 330 897 (HITACHI LTD) 6 September 1989 see column 4, line 4 - line 20 see column 6, line 57 - column 7, line 23; figures 1-4 ---	1,2,7
A	MESURES , no. 663, March 1994, PARIS FR , pages 38 40 42- , XP000434262 TUCKER: "acquisition de donnees - comment optimiser les echanges de donnees" see the whole document ---	1,2,7

-/-



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- *&* document member of the same patent family

Date of the actual completion of the international search

7 February 1997

Date of mailing of the international search report

21.02.97

Name and address of the ISA

Patent Office of the Netherlands

NL - 2500 AA Utrecht

Postbus 1000

Postbus 1000

Author and officer

Patent Office of the Netherlands

INTERNATIONAL SEARCH REPORT

Inter- national Application No

PC: / CH 96/03634

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>CONTROL AND INSTRUMENTATION, vol. 26, no. 3, March 1994, LONDON GB, pages 23-24 26, XP000446001 "data acquisiton" see figure page 23 -----</p>	1,2,7

INTERNATIONAL SEARCH REPORT

information on patent family members

Inte nna Application No

000/00 00 00 000

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A-294524	14-12-88	NONE	
EP-A-330897	06-09-89	JP-A- 1224657	07-09-89